**Sampling Techniques**

Presenter – Anil Koparkar

Moderator – Bharambe sir

Date:- 16/09/10

|  |
| --- |
|  |
|  |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | **Framework**   * Introduction * Need and advantages * Methods of sampling * Probability sampling * Simple Random Sampling – With & Without Replacement * Stratified Random Sampling * Systematic Random Sampling * Cluster Sampling * Non probability sampling * Convenient sampling * Purposive sampling   **Introduction**  Sampling is used most frequently in surveys. A purpose of a sample survey is to obtain information about population. Who *is* the target group for the study is called the **study population**. Who *in* the target group should be surveyed is called the **sample**. How *many* people should be surveyed is called the **sample size** and how should the people to be surveyed is the **sampling methods/ techniques**.  **Need and advantages of sampling techniques**  Census requires enormous time, trained personnel, money etc. and slightest bias can get magnified when no. of observations are increased. Also most of the time it is not feasible to collect data from whole population. Thus with sampling techniques data collected and analyzed in much less time and money. Moreover it is possible to obtain sufficiently accurate result by studying only a part of population. Thus reduces inspection fatigue.  **What is Sampling?**  It is a Procedure by which some members of a population are selected as representative of the entire population. Or it can also be defined as Procedure of selection of part of an aggregate to represent the whole population. Here sub-group thus selected to represent the whole population is known as SAMPLE. Thus sampling method is the scientific & objective procedure of selecting unit from a population and provides a sample that is expected to be representative of the population as a whole.    **Methods Of  Sampling**  Several methods are used to ascertain a particular aspect of the population, through an unbiased sample drawn from the population. Sampling is divided in two categories   1. **Probability sampling**   It is any method of sampling that utilizes some form of *random selection.* The procedure should assure that the different units in the population have some pre-estimated probabilities of being chosen. Probability sampling includes  •          Simple Random Method  •          Stratified Sampling  •          Systematic Sampling  •          Cluster Sampling  •          Multistage Sampling  •          Multiphase Sampling     1. **Non probability sampling**     It does not involve *random* selection. May or may not represent the population well. Used when researcher lacks a sampling frame for the population   Non-probability Sampling includes  •          Convenient sampling    •          Judgmental sampling  •          Quota Sampling    Lets see what are these methods one by one.  **Simple Random Sampling**    •          A sample selected such that each possible sample combination has equal probability of being chosen. Random does not mean haphazard.  This is applicable when population is small, homogeneous and readily available. E.g. patients in ward.  Two types of Simple Random Sampling.    1 ) **Simple random sampling without replacement**  **Definition**: It is a method of sampling such that every one of the possible sample of size *n* from N has the same probability1.  Theorem related to SRS without replacement  In simple random sampling without replacement   1. The probability of drawing a specified unit at the rth draw is equal to the probability of drawing it at the 1st draw. 2. The probability of including a specified unit in a sample of size *n is*   *n/*N. (N=total population, *n* =is sample size).1   •          In this method the population elements can enter the sample only once  •          The units once selected is not returned to the population in the next draw  Sampling design:  P(ir/i1,i2, i3,…….,ir-1)=1/N-r+1  = Pir/1-Pi1-Pi2-…..Pir-1  2 ) **Simple random sampling with replacement**    •           The population units may enter the sample more than once   This method is used when there is follow up in any survey. E.g School health survey.  Thus sampling design will be   ir/i1,i2,i3,….ir-1) = 1  P(i1,i2,i3,………,ir-1) = 1/N = Pir  Methods of selection of a simple random sampling:     1. **Lottery Method**   Ex. If we want to select 20 patients from 50 patients by lottery method, then –   1. All 50 patients can be given number serially from 1 to 50 on pieces of papers. 2. All these papers are folded and shuffled. 3. Draw one of it and note that no. 4. Now if we draw another piece of paper without replacing previous one – its SRS without replacement and 5. If we draw another piece of paper after replacing previous one – its SRS with replacement. 6. **Random number procedures** 7. By computer (Demo) 8. By random number table (Demo)   Example:- nine blocks in a certain administrative zone contain 793, 170, 970, 657, 1721, 864, 383 and 826 households respectively. If we want to select 6 households using a method of SRS Without Replacement.  [XLS](../SRSWR.xlsx)  Solution- the total no of households in all the nine blocks is 7987. The 1st step in selection of a random sample of households is to assume this serially numbered from 1 to 7987, by taking successive cumulative totals:793, 963, 1933, 2590, 4311, 5914, 6778, 7161, 7987. The 793 households in block 1 being assume to have the serial numbers 1 through 793, the 170 households in block 2 being assume to have the serial numbers 794 through 963 , and so on…  A reference to the 4 digit random numbers obtained by reading together columns by 9 through 12 in Random number Table.  Then give the following sample of households with serial numbers 7358, 922, 4112, 3596, 633 & 3999.  The corresponding households will be no. 197 from block 9, no. 129 from block 2, no. 1522 & no. 1006 from block 5, no. 633 from 1, and another number 1409 from block 5. It is obviously not necessary to have all the 7987 households in the 9 blocks numbered serially. Only the households in blocks 1, 2, 5 and 9 which happened to contain the random sample of 6 households have to be serially numbered in this example.  Advantages of Simple random sampling   * It is simple technique * Gives equal chances of selection for every individual from population.   Disadvantages of Simple random sampling   * If study is repeated, same samples couldn’t be identified * If population is heterogeneous, then SRS is not good technique to draw samples. * If population is divided in different strata, then SRS may leave representatives from some strata.   Thus to overcome problem of heterogeneity.  **Stratified Random Sampling**  Also sometimes called *proportional* or *quota* random sampling. If population is heterogeneous, in simple random sampling there is no guarantee that all the segments of population will be represented by population. Thus to increase precision apart from increase sample size, stratifying the population into more homogeneous group than entire population and drawing predetermined sample size from those strata is other way out. This method is called as Stratified Random Sampling.  Ex. To estimate the average income per household, it may be appropriate to group households into 2 or more (strata) according to rent paid by the household. Now the households in any stratum so formed are more homogeneous as compare to whole population.  Thus, the estimated income per household based on the stratified sample is likely to be more precise than that based on a simple random sample of same size drawn from the whole population.  Why you might prefer stratified sampling over simple random sampling.   1. It assures that you will be able to represent not only the overall population, but also key subgroups of the population, especially small minority groups. If you want to be able to talk about subgroups, this may be the only way to effectively assure you'll be able to. If the subgroup is extremely small, you can use different sampling fractions (f) within the different strata to randomly over-sample the small group.   When we use the same sampling fraction within strata we are conducting ***proportionate* stratified random sampling**. When we use different sampling fractions in the strata, we call this ***disproportionate* stratified random sampling**.   1. Second, stratified random sampling will generally have more statistical precision than simple random sampling. This will only be true if the strata or groups are homogeneous. If they are, we expect that the variability within-groups is lower than the variability for the population as a whole. Stratified sampling capitalizes on that fact.   Ex. Sample registration system(SRS). National sample survey (NSS)  **Systematic Random Sampling**    The method of sampling in which only the 1st unit is selected at random, the rest being selected automatically according to a predetermined pattern.  steps to follow in order to achieve a **systematic random sample**:   * number the units in the population from 1 to N (Say N = 100) * Decide on the n (sample size) that you want or need. (Say n=20) * k = N/n = the sampling interval size/sampling fraction (100/20 = 5) * Randomly select an integer (random start number) between 1 to k (i. e. 1 to 5. Say randomly selected r=4) * Then take every kth unit (every 5th number). i.e. nth sample= r+(n-1)k * Then samples selected will be  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 |  |  |  |  |  |  | | --- | --- | --- | --- | --- | | 21 | 22 | 23 | 24 | 25 | | 46 | 47 | 48 | 49 | 50 | | 71 | 72 | 73 | 74 | 75 | | 96 | 97 | 98 | 99 | 100 | |  |  |  |  |  |   So samples selected here are those colored in red..  When to do Systematic random sampling?  -population is large but homogeneous  -when complete list of population from which a sample is to be drawn is available  Merits of Systematic random sampling are  1. Procedure is simple and convenient to use.  2. Relatively less labor and time is needed.  3. If the population is sufficiently large and homogeneous and if numbering of subjects are available, then this method can provide good results.  4. If start number (1st sample) is known, whole samples can be determined. Thus sampling can be repeated or checked.  Demerits of Systematic random sampling  1. If engulfed in any hidden cyclic trend, then this technique can be dangerous. E.g. studying no of issue of books per day. If sampling interval (k)=7, then it will induce error.  **Cluster Sampling**    When we have to sample a population that's distributed across a wide geographic region we will have to cover a lot of ground geographically in order to get to each of the units we sampled. If we want to take a simple random sample of all the residents of India in order to conduct personal interviews. By SRS we will end up with respondents who come from all over the state. Our interviewers are going to have a lot of traveling to do. It is for perticularly this problem that **cluster or area random sampling** was invented.   In cluster sampling, we follow these steps:   * divide population into clusters (usually along geographic boundaries) * randomly sample clusters * measure all units within sampled clusters     If we have to do a survey of town, governments that will require us going to the towns personally. If we do simple random samples state-wide we'll have to cover the entire state geographically. Instead, we decide to do a cluster sampling of 10 states. Once these are selected, we go to *all 10 state* governments. Clearly this strategy will help us to economize on our mileage. Cluster or area sampling, then, is useful in situations like this, and is done primarily for efficiency of administration.    Merit of cluster sampling  1. Simple and time saving.  2. Better field supervision.  3. No sampling frame is required.  As the cluster size increases the cost decreases.  Demerit  Costlier and higher standard error.  Population proportionate to size (PPS) ex. DLHS-2  **Multi- Stage Sampling**  When sampling procedure is carried out in several stages using random sampling techniques, then it is called as Multi-stage sampling.  This procedure is used in large scale country wise or region wise surveys.  •     First stage- preparation of large sized sampling units  •     Randomly selecting a certain number  •     Second stage- Another list prepared from them  •     Sub-samples drawn by random sampling and so on….  Ex. Stages for sampling at national level are:   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Stages | 1 | 2 | 3 | 4 | 5 | 6 | |  | Country | State | District | Taluka | villages | Households/individuals |    Merits :-  1. Flexible and allows different selection procedure.  2. large no of units can by sampled for a given cost.  **Multi –Phase Sampling**    This technique is used to obtain supplementary information. Certain items of information collected from all units of sample and Other items collected from only some of sampling units (sub-sample)  Ex. In health examination survey among school children will be surveys in 1st phase. In 2nd phase for blood examination or X-ray of bone for maturity. Thus sub-sample will become smaller and smaller.  Advantage.  By such procedure, survey will be less costly, less laborious and more purposeful.      **NON- PROBABILITY SAMPLING**    These are the techniques which do not provide any basis for estimating the probability of item in the population for getting included in the sample. Thus there is no assurance that every element has some specifiable chance of being included. Here judgment of researcher / organizer plays great role. Thus personnel error has great chances to enter.  Other characters-–   * Representativeness is in question as sampling error cannot be measured * More suitable for small in-depth enquiries than large surveys. * Saves time and money (speed and administration convenience) * More flexible   Types of Non-probability sampling   1. **Convenient/ haphazard Sampling**   In this technique samples are selected at the convenience of the researcher. Thus it is useful in formulative or explorative studies, pilot surveys, testing questionnaire, ..etc.  Ex. 1. Choosing fruits from basket, pilot testing of thesis questionnaire, The "person on the street" interviews conducted frequently by television news programs. Sampling those most convenient. Gets a quick reading of public opinion.  In technique not to be used in descriptive or diagnostic studies or for causal studies.   1. **Purposive or deliberate sampling** 2. **Judgment sampling** –   in this method researcher Purposively or deliberately draws a sample which he/she thinks is representative. Thus personnel biases of investigator have great chances.  Ex. Communication partner, selecting talkative children for interviewing to find the cause of tobacco chewing. Selecting mothers who are willing to participate in study of Health Care Seeking for Newborn Danger Signs.   1. **Quota sampling**   In this method selection is based on some basic parameters like age, sex, income etc… of population so as to make it representative. In this method field workers are assigned quotas of number of units satisfying the required characteristics for collecting data. (looks similar to stratified sampling, but differ in fact that discretion of field worker is not found in stratified sampling which makes random sample from each cell.  Properties – when parameters are large, the number of cells increases. It misleads if relevant parameter is omitted. In this method field worker tend to visit respondents who are more likely to be available and accessible.  Limitation – its often difficult to get accurate and up-to-date proportion of respondent assigned to each cell.  Ex. Communication behavior of user to be carry out on a quota sample from a population having the following parameters:   1. %age of grad., post. Grad.& doctorates are 20, 35, 40 resp. 2. Male:female = 60:40 3. If sample size is 200, find the quota sampling as per above 2 parameters.   Ans:- Qualification ratio/ proportion G:P:D = 4:7:9, gender ratio/ proportion M:F = 3:2  Therefore graduate male = 4/20 × 60 = 12 or 3/5 × 20 = 12.  Likewise all cells can be filled and then samples are collected purposively.  Distribution of particulars of population in %   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Sex\Qln | Grad | Post grad. | Doct | Total | | Male | 12 | 21 | 27 | 60 | | Female | 8 | 14 | 18 | 40 | | Total | 20 | 35 | 45 | 100 |   Quota sample in numbers as per parameters   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Sex\Qln | Grad | Post grad. | Doct | Total | | Male | 12 | 21 | 27 | 60 | | Female | 8 | 14 | 18 | 40 | | Total | 20 | 35 | 45 | 100 |         . | |

**References:**

1. Sukhatme PV, Sukhatme S, Sukhatme BV, Asok C. Sampling theory of surveys with applications. Iowa state university press, Ames, Iowa, USA. 3rd ed. 1997. 1-182.
2. Lwanga SK, Cho YT, Ayeni o. Teaching Health Statistics Lesson And Seminar Outlines. WHO Geneva, 1999. 66-78.
3. Sampling Methods in Medical Research. Available online at URL: <http://www.similima.com/thesis62.html>.
4. Rao KV. Biostatistics. A manual of statistical methods for use in health, nutrition = and anthropology. 2nd ed. Jaypee brothers medical publisher (P) ltd. 2007. 11-18.
5. Sridhar MS. Research methodology; part 5 – sampling and sampling strategy. Available online at URL: <http://www.scribd.com/doc/203905/Research-Methodology-Part-5-Sampling-Sampling-Strategy-or-Plan#fullscreen:off>
6. Amol R Dongre, Pradeep R Deshmukh and Bishan S Garg. Awareness and Health Care Seeking for Newborn Danger Signs Among Mothers in Peri-Urban Wardha. Indian Journal of Pediatrics. May 9, 2008. 2
7. Dongre AR, Deshmukh PR, Murali N, Garg BS. Tobacco consumption among adolescents in rural Wardha: Where and how tobacco control should focus its attention?. Indian Journal of Cancer ; Sept 2008: 45 (3). 101